CGIAR Research Program on
Climate Change, Agriculture and Food Security (CCAFS)

Data Management Plan
Video Transcript

October 2013

These materials were produced for and with funding from the Climate Change Agriculture and Food Security Research Program of the Consultative Group on International Agricultural Research (CGIAR).

Creating a Data Management Plan

This short presentation accompanies the document and example plan in this section of the Data Management Support Pack. We will consider the overarching policies for data management and the detailed procedures.

Introduction

The aim of a data management plan is to help you produce high quality data from your research leading to results that are robust and reliable.

By considering the data management tasks needed from the start of the project you can ensure you have the necessary resources available. In your project plans you must allow enough time for team members to carry out the tasks assigned to them. Consider the skills available in your team – do you need more staff or do you need training for existing staff? What about equipment? This could be network storage devices for shared storage and backups; or you might need hand-held devices for collecting data. Be sure to allocate sufficient financial resources to data management – do you have a line in your project budget for data management?

Policy & Plan

In planning for data management we consider two levels – at the project level we will refer to a data management "Policy" – this policy sets out the main principles and agreements. For example you might have a general agreement that all data generated by the project will be put into the public domain within 2 years of collection. The principal investigator would be the person responsible for making sure there is a project data management policy.

At the project activity level we consider the data management plan – this should detail the procedures needed to carry out the policy. For example the plan would detail how we are going to archive the data – the data will be archived using Dataverse and the archive will include a data quality report, a metadata document, original questionnaires, etc. The plan would generally be drawn-up and implemented by the data manager or whoever has been given overall data management responsibilities in the project.

Project Level Policy

One of the first questions a PI should address is "Do I need a Data Manager?" The choice is basically between having a specialist data manager who takes responsibility for all data management tasks and allocating these responsibilities to scientists within the team. This is a debateable point and the decision depends on the size and complexity of the study and the skills of the scientists. We would recommend the inclusion of a data manager in most projects – from our experience we have found that scientists have too much other work to do and without a dedicated data manager the necessary tasks are at best rushed or even skipped all together.

We'll mention here some of the headings for overarching principles and agreements.

We'll start with data ownership which is often a contentious issue. Thus it is important to draw up agreements from the outset to avoid problems later. Data ownership is covered in more detail in a separate guide.

Data sharing and access is linked to data ownership in many ways. Team members must have access to the data they need for their work and we would recommend using shared data and document stores. However, you do need to consider confidentiality so it might be that data access occurs in a staged manner with more team members being given access with time. Data ownership and sharing agreements can help avoid problems such as a researcher disappearing with the only copy of the data.

If the research involves collecting data on individuals the researcher would generally need to obtain ethical approval or establish the code by which the project will work. Respondents must be fully informed of the purpose of the study and personal data must remain confidential. If any copyrighted data or tools are used then permission must be sought from the copyright holder. For example in one project a particular copyrighted tool was used and we were given permission to use it and the resulting data on condition that no changes were made to the tool. However, one country team made changes and permission was subsequently withdrawn; we had to discard the data from that country.

At this level there is likely to be a broad statement saying something like all data generated by the project will be put into the public domain within say 2 years.

Quality standards and security covers aspects such as data checking and correction and regular backups. At this level we could say for example “All project data will be subject to a quality assurance process. Regular backups will be taken throughout the project”.

Activity Level Plan

At the activity level you should add the detail and here we’ll go through a few pointers on the sort of things you should be including in your plan.

For data collection what capture methods will you be using? Will this be a survey or experiment? Describe the methods to be used. What technology will you be using – e.g. paper questionnaires, mobile devices, etc.

Briefly describe the data you intend to collect – for example we will be collecting data on 20 households in each of 7 villages.

Are you likely to be using any secondary data, if so then how will you source this data? Who owns it?

When it comes to computerising your data think how you are going to enter it – what software will you be using; will you have a customised data entry system; if recording directly onto hand-held devices what mechanisms do you have to ensure data quality – e.g. two researchers will be present during data collection for validation purposes. Have data entry staff been trained? Are you using double data entry?

How do you intend to ensure your data is of high quality? Detail the check you intend to carry out – for example 12yr olds cannot be grandparents, harvest date cannot be before planting date.

Describe the structure of the data – in particular how many levels of data you are expecting and what they are – for example household level, individual level – how many cases are expected at each level? How will you link data at different levels? What formats are to be used for storing the data?

Storage and backup

Will you be using a data and document storage facility – if yes, then describe the system to be used. We describe data and document stores in more detail in documents and videos in a separate section of this pack. Describe your backup procedures – how often will you take backups; who is responsible for making sure they are done; where will they be stored (e.g. locally on external hard drives or in cloud storage)

Legal Aspects

Do you have a consent form for respondents? This should be included in your archive. Do you need ethical approval and where is this being sought? What mechanisms do you have for ensuring confidentiality of personal data?

Have all team members been made aware and agreed to the terms of the Project Data Ownership Agreement?

Are you using any copyrighted materials and if so have you sought permission to use them? Detail any legal restrictions that might impact on how the data are used. We mentioned earlier the case where a country team made wording changes to a data collection instrument which meant we were not able to use the data from that country.

Archiving and Preservation

The decision on whether or not to archive would generally be taken at the project level but the decision on where to archive might be taken at the activity level. Where do you intend to archive your data? Will there be any access restrictions on the archive? Please see the separate section on Archiving for more information on this topic.

Training & Responsibilities

If you don't already have the required skills within the team then new staff will need to be brought on board or existing staff will need to be trained. Regardless of the team composition you will often need to train enumerators and data entry staff so detail how you intend to carry out the necessary training and who will be responsible for this.

Summary

So to summarise what we are suggesting is a project policy to define principles and agreements – think of this as "what you intend to do".

We then have activity plans detailing the intended procedures – this can be thought of as "how you plan to do it".

Towards the end of the project or activity it is a good idea to produce a Data Management Report - this would detail "what you actually did and how well it worked – or what didn't work".

Appendix I – CCAFS Data Management Support Pack

This document is part of the CCAFS Data Management Support Pack produced by the Statistical Services Centre, University of Reading, UK. The following materials are available in the pack:

0. Data Management Strategy
    a. CCAFS Data Management Strategy
1. Research Protocols
    a. Writing Research Protocols – a statistical perspective
    b. Preparation of Research Protocols – Good Practice Case Study
    c. What is a Research Protocol, and how to use one (Video & Transcript)
    d. Details of what a Research Protocol should contain (Video & Transcript)
2. Data Management Policies & Plans
    a. Creating a Data Management Plan
    b. Data Management Plan (Video & Transcript)
    c. Example Data Management Activity Plan
    d. Example Consent Form
3. Budgeting & Planning
    a. Budgeting & Planning for Data Management
    b. ToR Data Support Staff
    c. Budgeting & Planning (Video & Transcript)
4. Data Ownership
    a. Data Ownership and Authorship
    b. Template – Data Ownership Agreement
    c. CCAFS Data Ownership & Sharing Agreement
    d. Data Ownership & Authorship (Video & Transcript)
5. Data & Document Storage
    a. Creating and Using a DDS
    b. DDS Introduction – (Video & Transcript)
    c. DDS Organisation – (Video & Transcript)
    d. DDS Ownership – (Video & Transcript)
    e. Introduction to Dropbox – (Video & Transcript)
6. Archiving & Sharing
    a. Archiving & Sharing Data
    b. Data and Documents to Submit for Archiving – a checklist
    c. MetaData
    d. Archiving & Sharing (Video & Transcript)
    e. Metadata (Video & Transcript)
    f. CCAFS HBS Questionnaire
    g. CCAFS HHS Code Book
    h. CCAFS Training Manual for Field Supervisors

7. CCAFS Data Portals
    a. Portals for CCAFS Outputs
    b. AgTrials Summary
    c. CCAFS-Climate Summary
    d. DSpace Introduction
    e. Introduction to Dataverse (Video & Transcript)
    f. Creating a Dataverse (Video & Transcript)
    g. Dataverse Study Catalogue
    h. CCAFS Dataverse (Video & Transcript)
8. Data Quality & Organisation
    a. Data Quality Assurance
    b. Guidance for handling different types of Data
    c. Transition from Raw to Primary Data
    d. Data Quality Assurance (Video & Transcript)
    e. Guidance for handling different types of data (Video & Transcript)
    f. Transition from Raw to Primary Data (Video & Transcript)